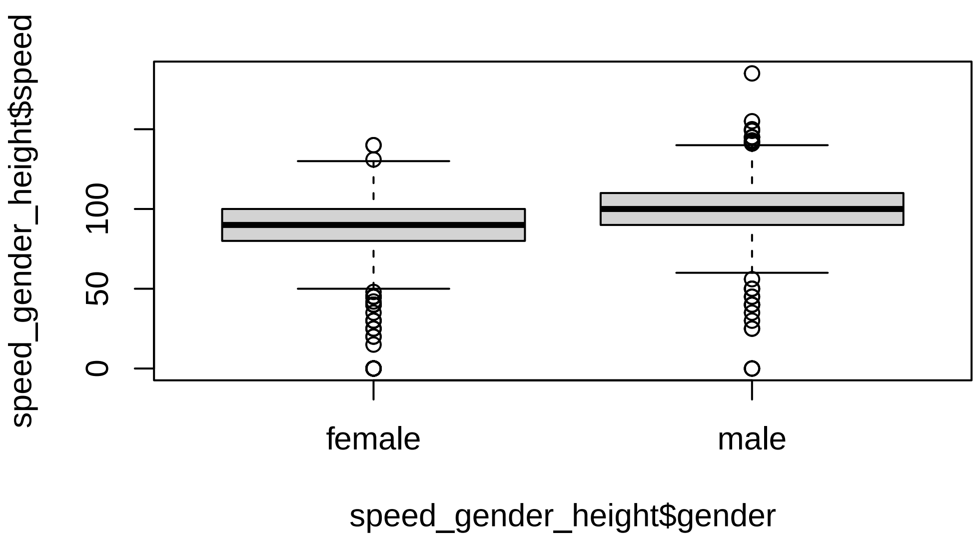
Homework 6

Mark Schulist

1.

```r
source("https://www.openintro.org/data/R/speed_gender_height.R")
```

```r
boxplot(speed_gender_height$speed ~ speed_gender_height$gender)
```



The boxplots appear mostly normal, especially given the large sample sizes for each group. There is some left skewedness, but it is not too bad.

$$H_0 : \mu_1 = \mu_2$$
$$H_a : \mu_1 \neq \mu_2$$

```r
t.test(speed_gender_height$speed ~ speed_gender_height$gender)
```

```
	Welch Two Sample t-test

data:  speed_gender_height$speed by speed_gender_height$gender
t = -8.326, df = 840.93, p-value = 3.372e-16
alternative hypothesis: true difference in means between group female and group male is not e
95 percent confidence interval:
 -13.385173  -8.278181
sample estimates:
mean in group female   mean in group male
            87.08651             97.91818
```

The p-value is extremely small, so we reject the null hypothesis and have evidence to show that there is a true different in means between male and female drivers' top speed driven.

$$H_0 : \tilde{\mu}_1 = \tilde{\mu}_2$$
$$H_a : \tilde{\mu}_1 \neq \tilde{\mu}_2$$

```r
wilcox.test(speed_gender_height$speed ~ speed_gender_height$gender)
```

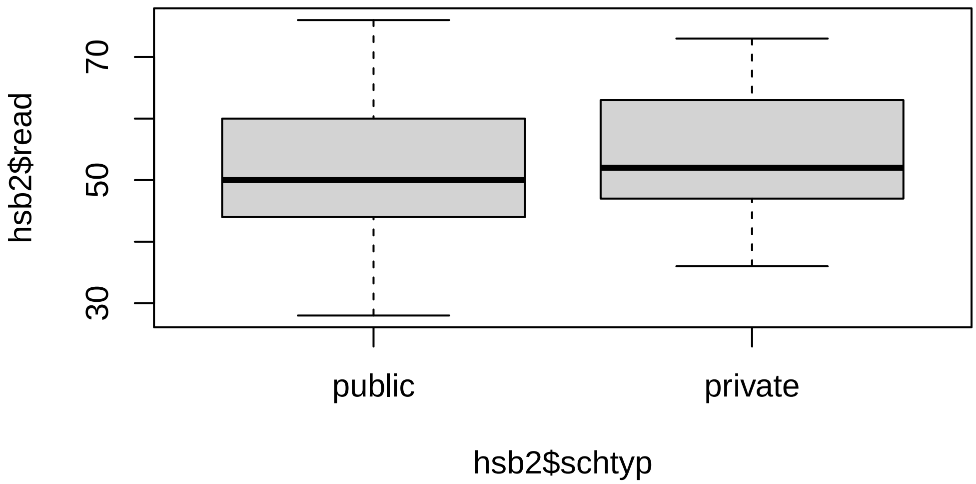
```
	Wilcoxon rank sum test with continuity correction

data:  speed_gender_height$speed by speed_gender_height$gender
W = 131012, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

The p-value is very small, so so we reject the null hypothesis and have evidence to show that there is a true different in medians between male and female drivers' top speed driven.

2.

```r
source("https://www.openintro.org/data/R/hsb2.R")
```

```r
boxplot(hsb2$read ~ hsb2$schtyp)
```



The boxplots appear mostly normal, although there is some right skew in both of them. Given we have 100 samples for each group, this should be fine.

$$H_0 : \mu_1 = \mu_2$$
$$H_a : \mu_1 \neq \mu_2$$

```
t.test(hsb2$read ~ hsb2$schtyp)
```

```

    Welch Two Sample t-test

data:  hsb2$read by hsb2$schtyp
t = -1.3258, df = 47.496, p-value = 0.1912
alternative hypothesis: true difference in means between group public and group private is no
95 percent confidence interval:
 -6.052601  1.243078
sample estimates:
 mean in group public mean in group private
             51.84524              54.25000
```

The p-value is above 0.05, so we do not reject the null hypothesis. We do not have evidence to conclude that the true difference in mean reading scores between public and private schools is different.

$$H_0 : \tilde{\mu}_1 = \tilde{\mu}_2$$

$$H_a : \tilde{\mu}_1 \neq \tilde{\mu}_2$$

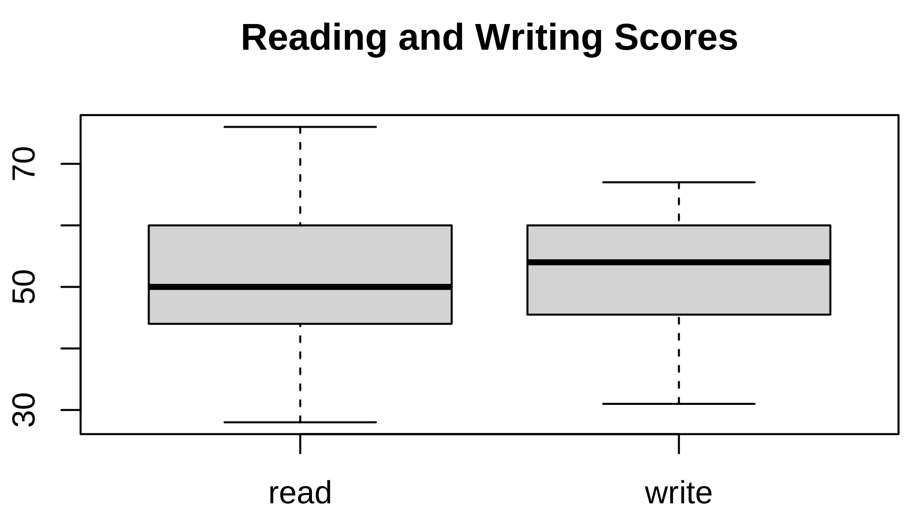
```
wilcox.test(hsb2$read ~ hsb2$schtyp)
```

```

	Wilcoxon rank sum test with continuity correction

data:  hsb2$read by hsb2$schtyp
W = 2313.5, p-value = 0.2113
alternative hypothesis: true location shift is not equal to 0
```

The p-value is above 0.05, so we do not reject the null hypothesis. We do not have evidence to conclude that the true difference in median reading scores between public and private schools is different.

3.

```
boxplot(hsb2$read,
        hsb2$write,
        names = c("read", "write"),
        main = "Reading and Writing Scores"
        )
```



The writing median appears to be higher, but overall the distributions look very similar. The writing scores are left skewed and the reading scores are (a little) right skewed.

Our samples are not independent—they are paired.

Because the reading and writing distributions are (mostly) normal, the assumption of normality is reasonable. There are also 100 samples, which is plenty.

$$H_0 : \mu_{\text{diff}} = 0$$
$$H_a : \mu_{\text{diff}} \neq 0$$

```
t.test(hsb2$read, hsb2$write, paired = T)
```

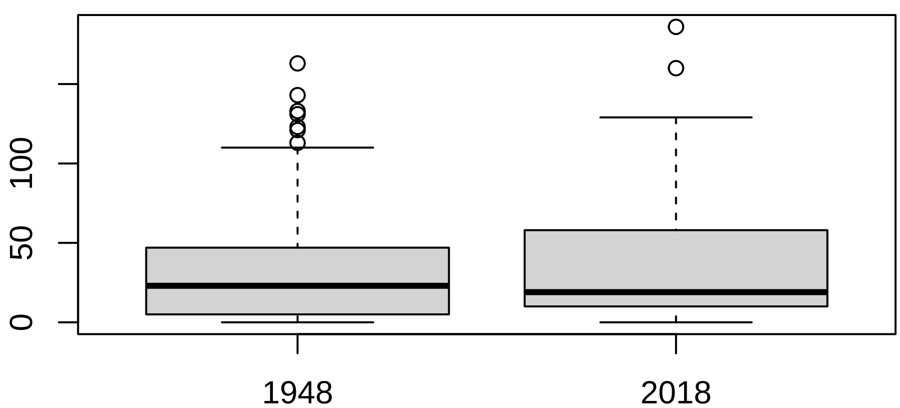
```

	Paired t-test

data:  hsb2$read and hsb2$write
t = -0.86731, df = 199, p-value = 0.3868
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 -1.7841424  0.6941424
sample estimates:
mean difference 
         -0.545 
```

The p-value is above 0.05, so we do not reject the null. We do not have evidence to suggest that the difference between each person's reading and writing scores is not equal to 0.

4.

```
source("https://www.openintro.org/data/R/climate70.R")
```

```
boxplot(climate70$dx90_1948, climate70$dx90_2018, names = c("1948", "2018"))
```



There is not a large difference between the distributions, although 2018 seems to have a longer right tail. These distributions are not normal, although we have enough samples (197) that the normality is probably not a big concern.

We have paired samples. They are from the same locations on different dates.

$$H_0 : \mu_{\text{diff}} = 0$$
$$H_a : \mu_{\text{diff}} \neq 0$$

```r
t.test(climate70$dx90_1948, climate70$dx90_2018, paired = T)
```

```
	Paired t-test

data:  climate70$dx90_1948 and climate70$dx90_2018
t = -2.3702, df = 196, p-value = 0.01875
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 -5.3101319 -0.4868224
sample estimates:
mean difference
      -2.898477
```

The p-value is below α, so we reject the null hypothesis. We have evidence that there is a difference in the average number of days above 90 degrees between 1948 and 2018.

5.

a. Paired, they come from the same days.

b. Paired, they come from the same items.

c. Not paired, the students are independent in both samples.

d. Not paired, the salaries are independent in both samples.

e. Paired, the people are the same in both samples, just at different points in time.